Image-Based Target Detection and Tracking

Aggelos K. Katsaggelos
Thrasyvoulos N. Pappas
Peshala V. Pahalawatta
C. Andrew Segall

SensIT, Santa Fe
January 16, 2002

Introduction

- Objective: Impact of visual sensors on benchmark and operational scenarios
- Project started June 15, 2001
- Video data acquisition
- Initial results with imaging/video sensors
  - For *Convoy Intelligence Scenario*
  - Detection, tracking, classification
  - Image/video communication

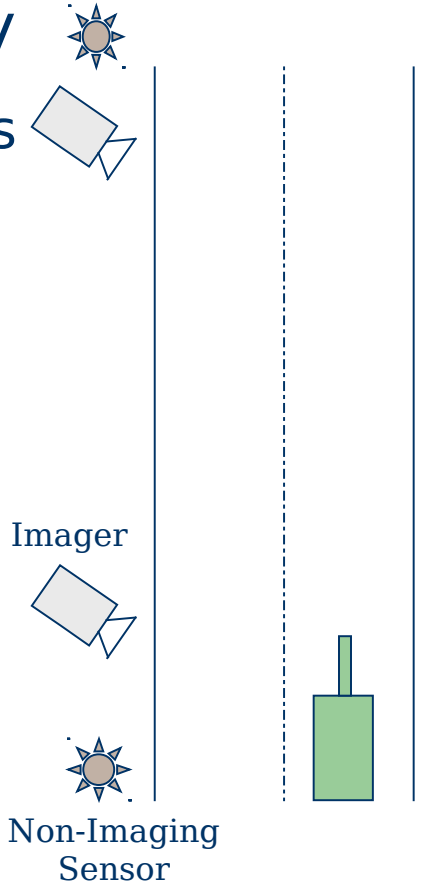
Battlefield Scenario*

Gathering Intelligence on a Convoy

- Multiple civilian and military vehicles
- Vehicles travel on the road
- Vehicles may travel in either direction
- Vehicles may accelerate or decelerate

Objectives

- Track, image, and classify enemy targets
- Distinguish civilian and military vehicles and civilians
- Conserve power



* Jim Reich, Xerox PARC

Experimental Setup

Imager Type

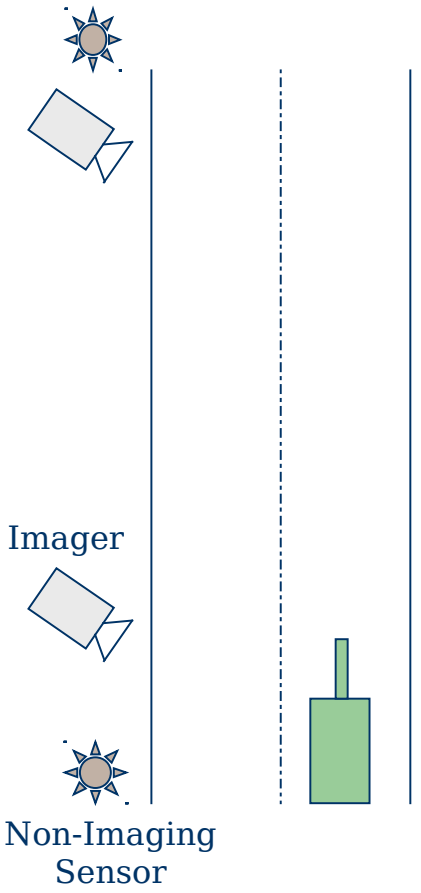
- 2 USB cameras attached to laptops (uncalibrated)
- Obtained grayscale video at 15 fps

Imager Placement

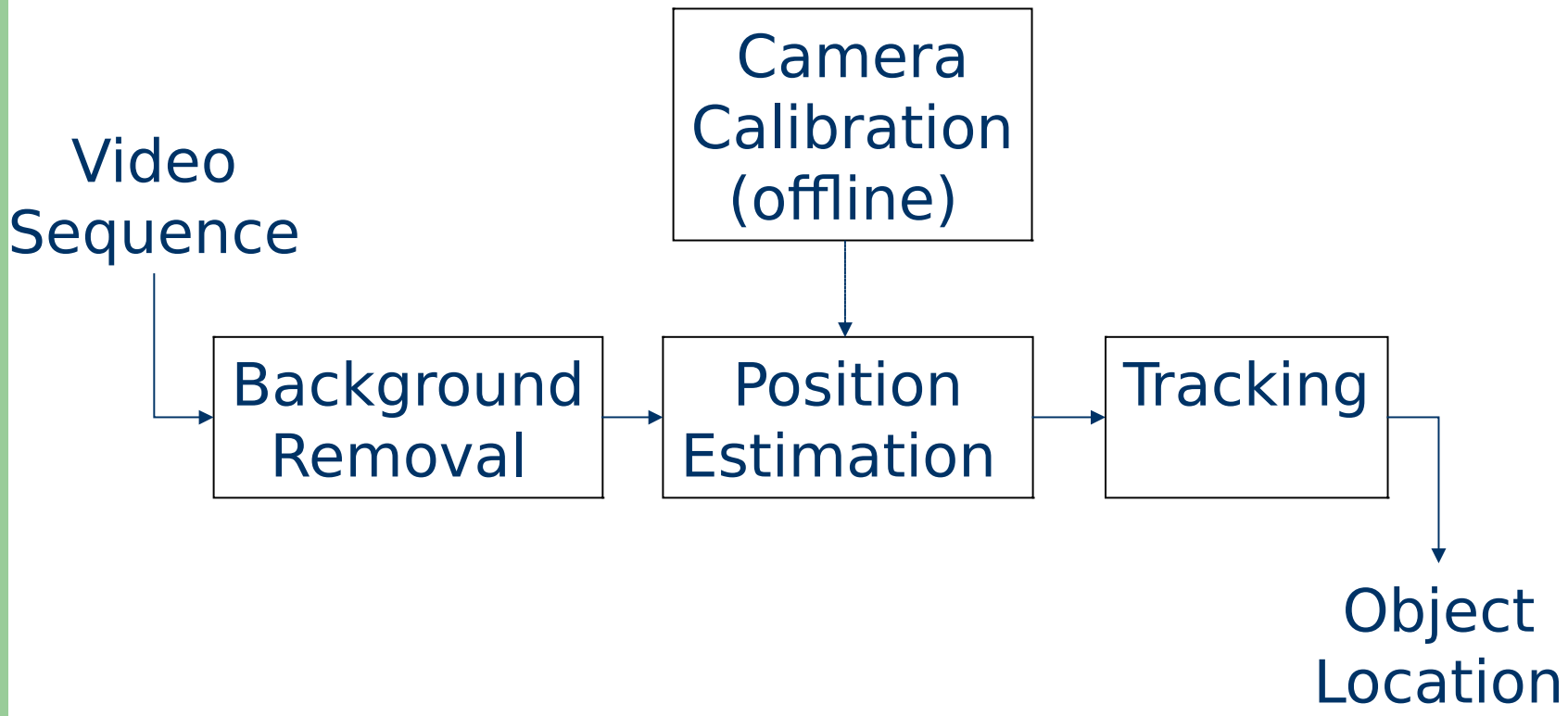
- 13 ft from center of road, 60 ft apart
- Cameras placed at an angle relative to the road to capture large field of view

Test Cases:

- One target at constant velocity of 20mph
- One target starts at 10mph, increases to 20mph
- One target starts at 10mph, stops and idles for 1min, and then accelerates
- Two targets from opposite directions at



Tracking System



Background Model *

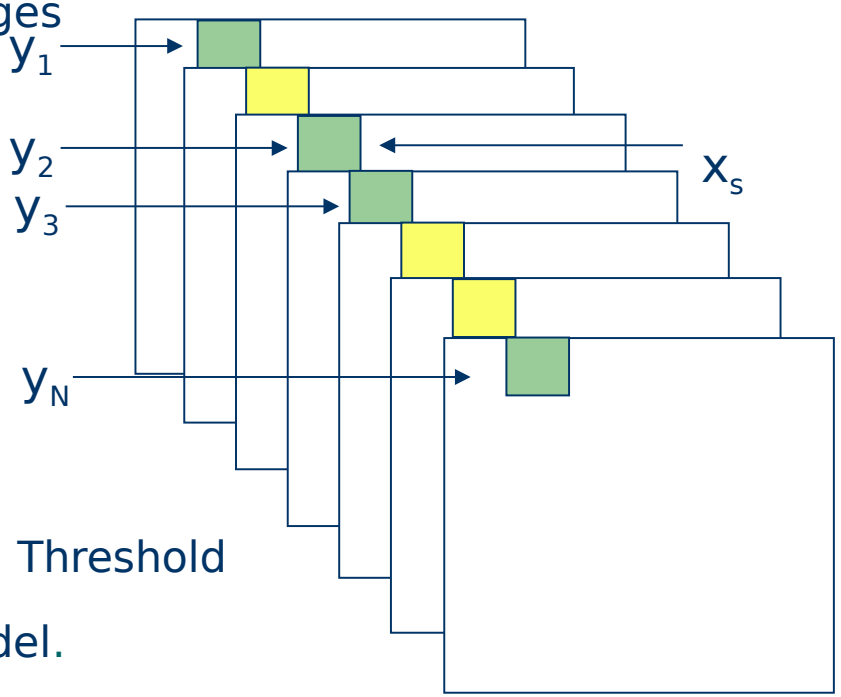
Basic Requirements

- Intensity distribution of background pixels can vary (sky, leaf, branch)
- Model must adapt quickly to changes

Basic Model

Let $\Pr(x_t)$ = Prob x_t is in background

$$\Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_t - y_i)^2}{2\sigma^2}}$$



- $y_i = x_s$, some $s < t$, $i = 1,2, \ldots, N$
- x_t is considered background if $\Pr(x_t) >$ Threshold
- Equivalent to a Gaussian mixture model.
- σ based on MAD of consecutive background pixels

* Ahmed Elgammal, David Harwood, Larry Davis "Non-parametric Model for Background Subtraction," *6th European Conference on Computer Vision*, Dublin, Ireland, June/July 2000.

Estimation of Variance (σ)

Sources of Variation

- Large changes in intensity due to movement of background (should not be included in σ)
- Intensity changes due to camera noise

Estimation Procedure

- Assume $y_i \sim N(\mu, \sigma^2)$
- Then, $(y_i - y_{i-1}) \sim N(0, 2\sigma^2)$
- Find Median Absolute Deviation (MAD) of consecutive y_i's

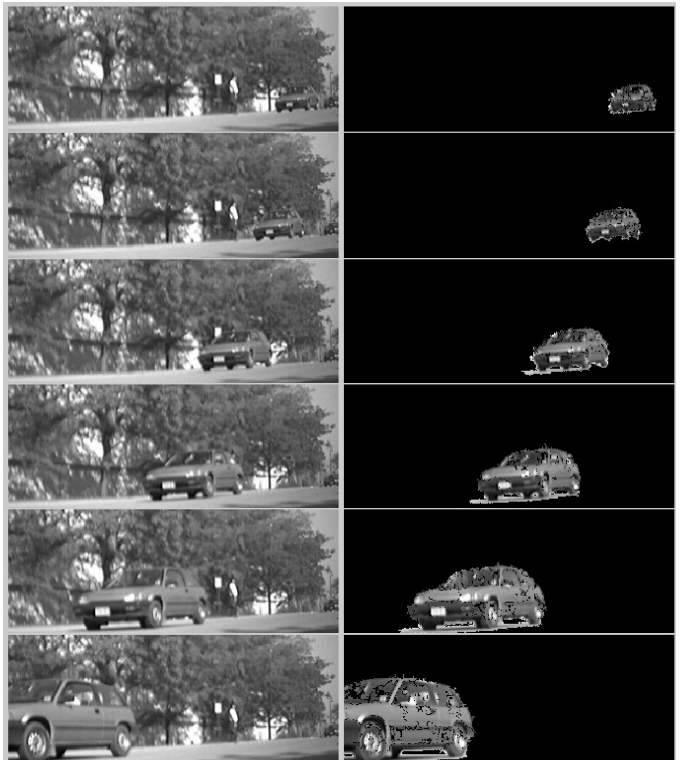
$$m = median(|y_i - y_{i-1}|)$$

- Use m to find σ from:

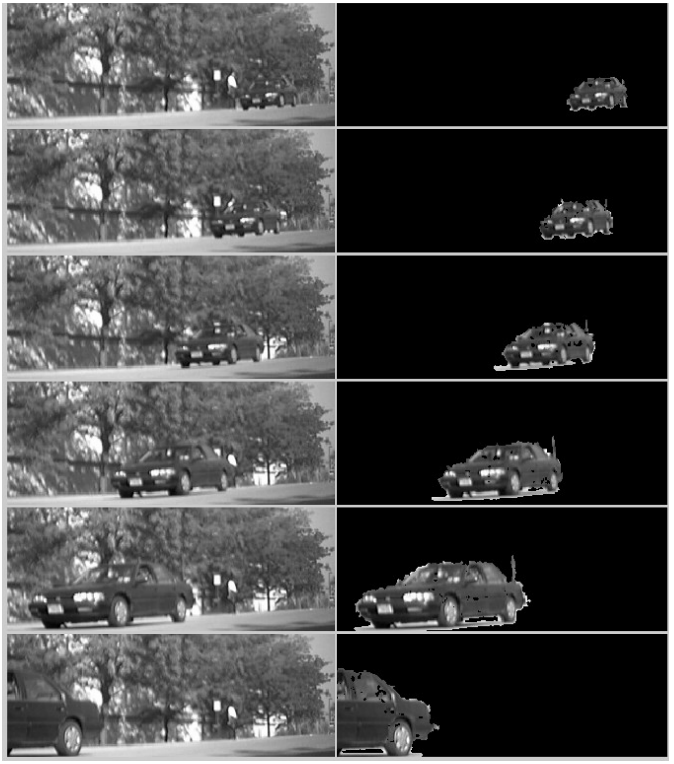
$$\sigma = \frac{m}{0.68\sqrt{2}}$$

Segmentation Results

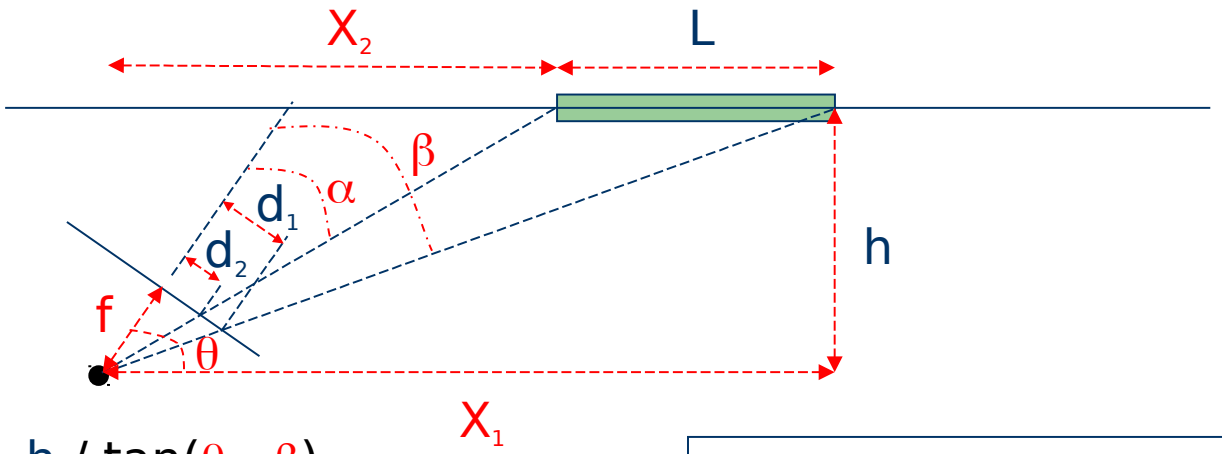
Foreground extraction of first target at 20mph



Foreground extraction of second target at 20mph



Camera Calibration



1. $X_1 = h / \tan(\theta - \beta)$

2. $X_2 = h / \tan(\theta - \alpha)$

3. $L = X_1 - X_2$
$= h [1 / \tan(\theta - \beta) - 1 / \tan(\theta - \alpha)]$

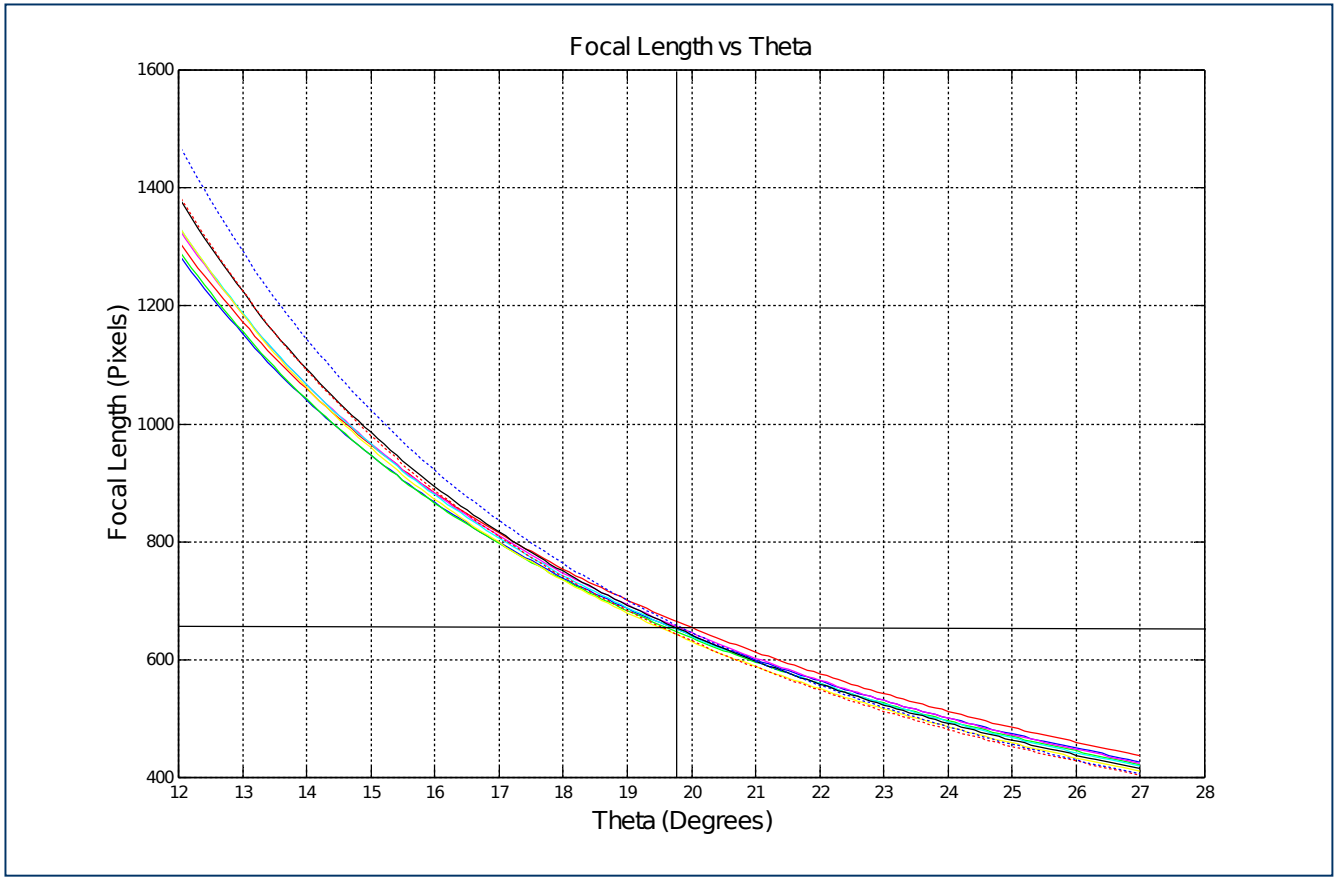
$$\frac{L}{h} = \frac{1 + \tan(\theta)\frac{d_1}{f}}{\tan(\theta) - \frac{d_1}{f}} - \frac{1 + \tan(\theta)\frac{d_2}{f}}{\tan(\theta) - \frac{d_2}{f}}$$

Variables to estimate: f and θ

Assumptions

- **Ideal pinhole camera model**
- **Image plane is perpendicular to road surface**

Calibration Results



Tracking

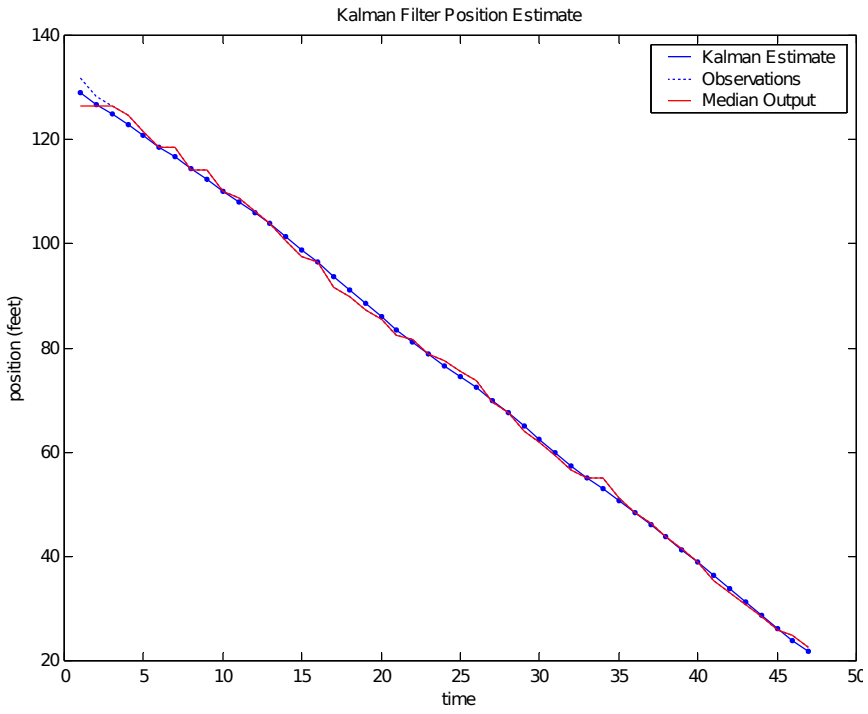
Median Filtering

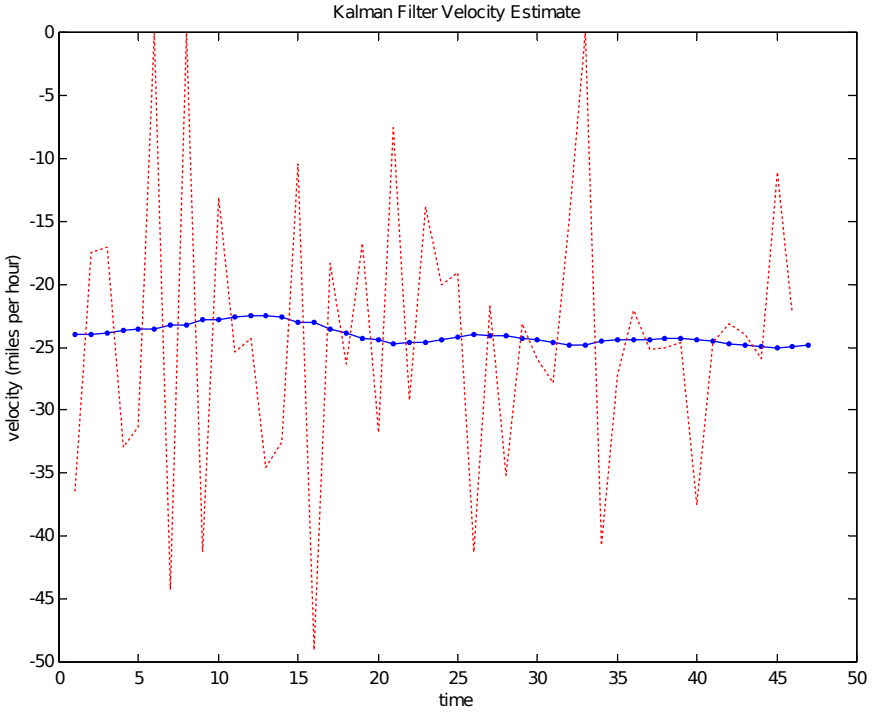
- Used to smooth spurious position data
- Doesn’t change non-spurious data

Kalman Filtering

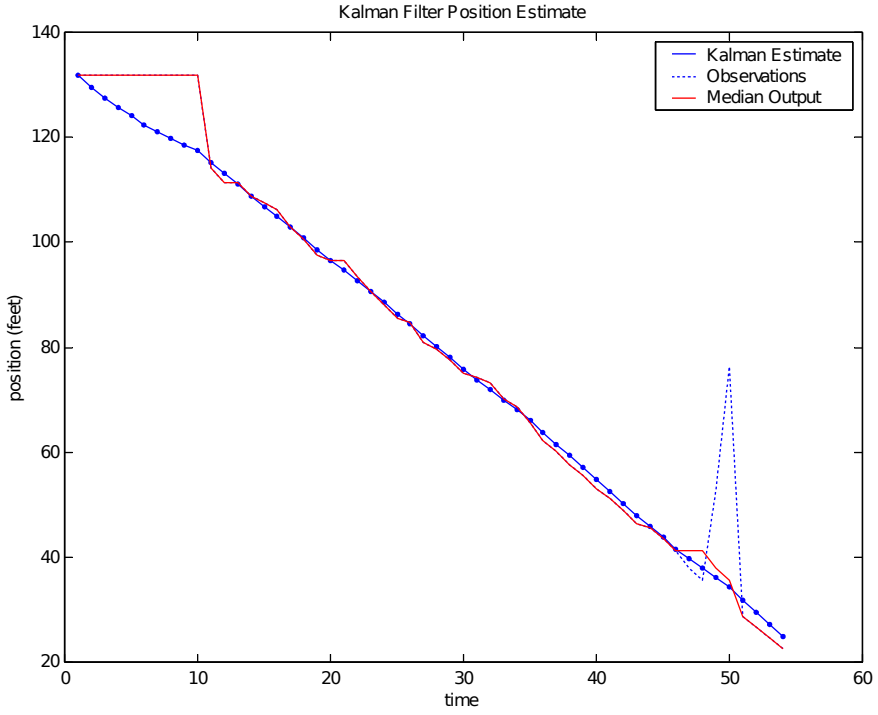
- Constant acceleration model
- Initial conditions set by our assumptions
- Used to track position and velocity

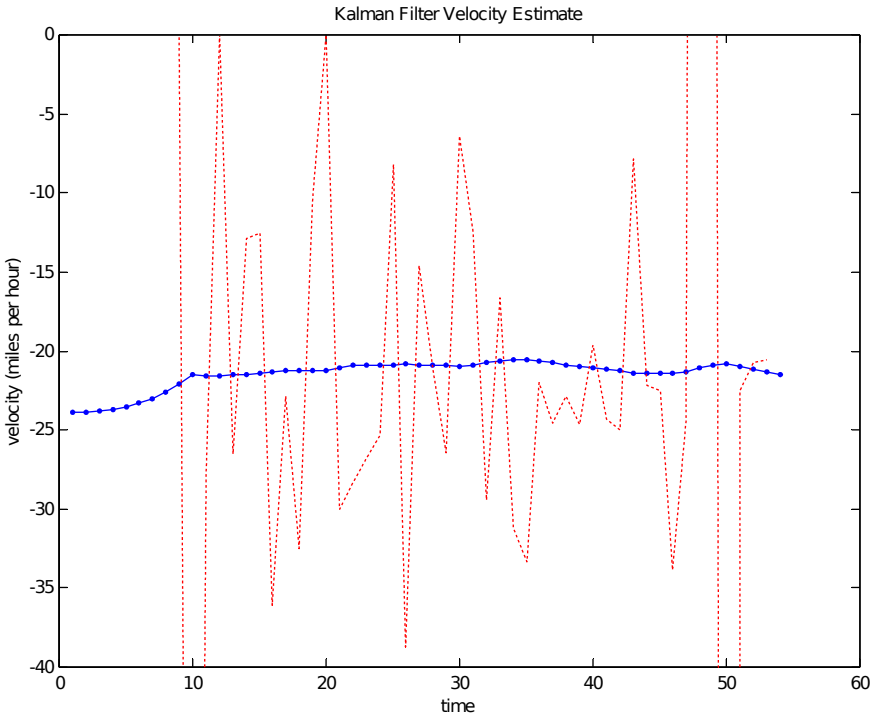
Results: Target #1 20 mph



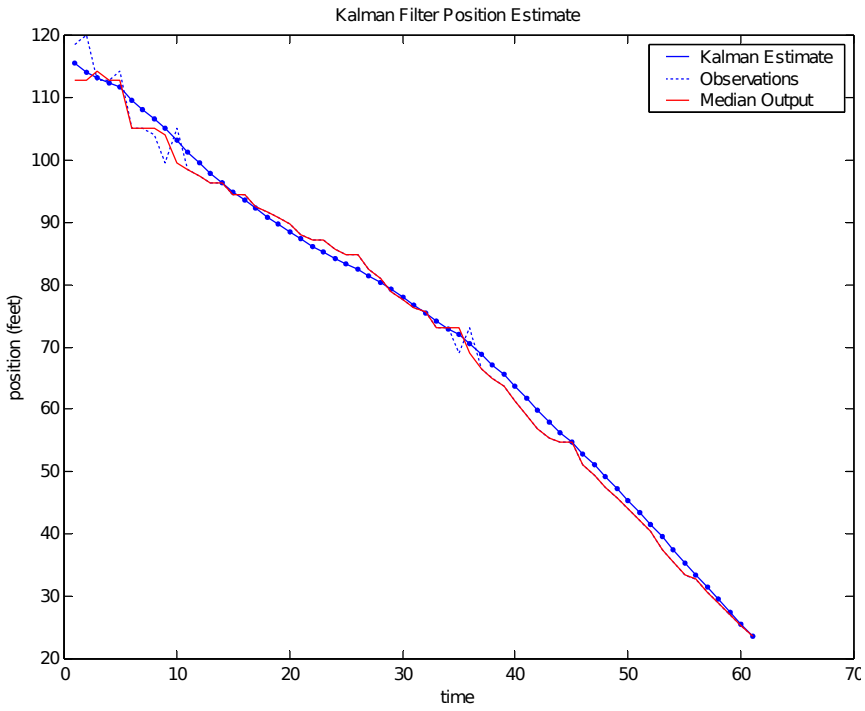


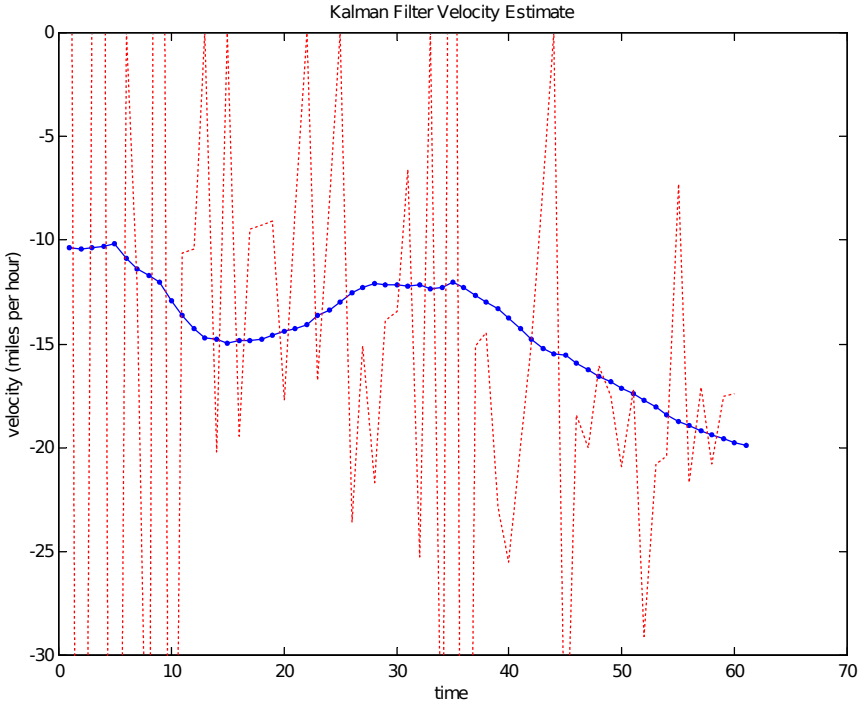
Results: Target #2 20 mph



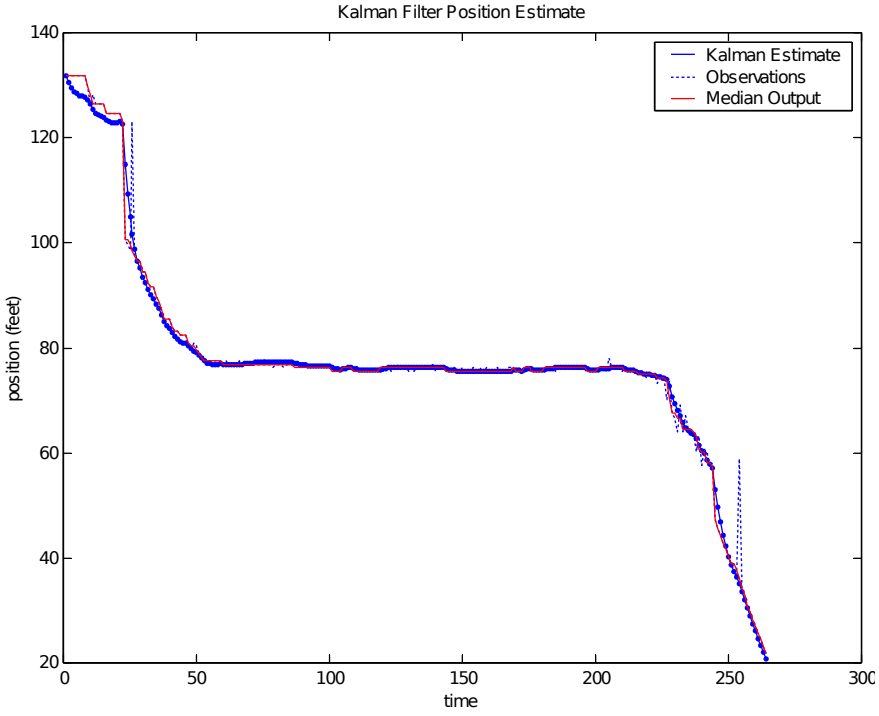


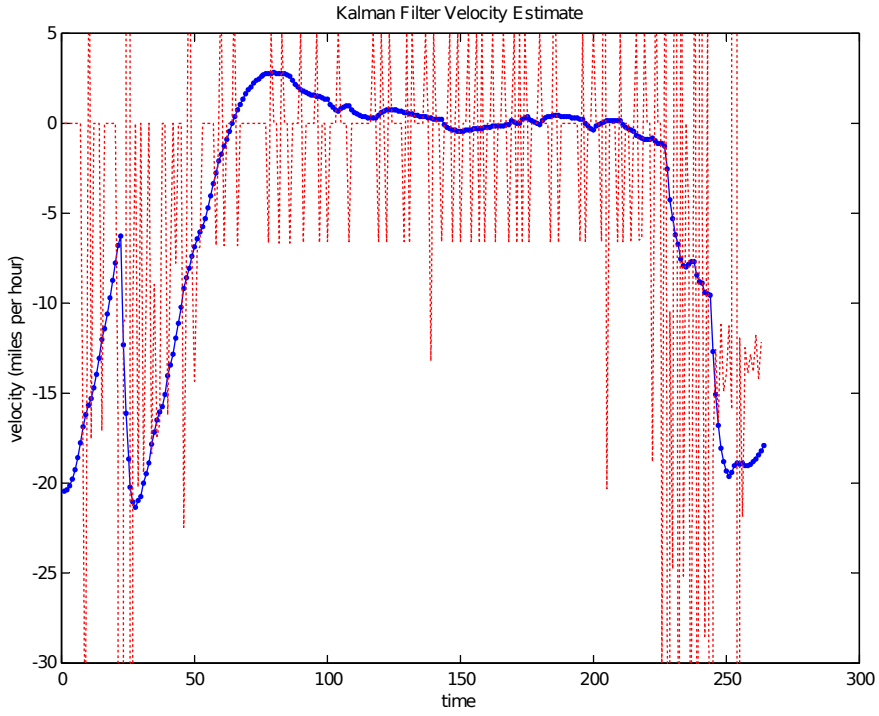
Results: Target #1 10-20 mph





Results: Target #2 Stop-Start





Work in Progress

- Improving and automating camera calibration process
- Improving foreground segmentation results using
  - background subtraction
  - image feature extraction (color, shape, texture)
  - spatial constraints in the form of MRFs
  - information from multiple cameras
- Estimating accuracy of segmentation
  - use result to improve Kalman filter model
- Multiple object detection
- Object recognition
- Integration with other sensors

Other Issues

- Communication between sensors
  - When/what to communicate
  - Power/delay/loss tradeoffs
- Communication of image/video
  - Error resilience/concealment
  - Low-power techniques
- Communication of data from multiple sensors
  - Multi-modal error resilience

Low-Energy Video Communication*

- Method for efficiently utilizing transmission energy in wireless video communication
- Jointly consider source coding and transmission power management
- Incorporate knowledge of the decoder concealment strategy and the channel state
- Approach can help prolong battery life and reduce interference between users in a wireless network

* C. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Transmission energy minimization in wireless video streaming applications," Proc. of Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, Nov. 4-7, 2001.

Image-Based Target Detection and Tracking

Aggelos K. Katsaggelos
Thrasyvoulos N. Pappas
Peshala V. Pahalawatta
C. Andrew Segall

SensIT, Santa Fe
January 16, 2002